APPENDIX I

STATISTICAL PROCEDURE TO DETERMINE IF A SITE-SPECIFIC OBJECTIVE FOR A CHEMICAL WITH A HIGH NATURAL BACKGROUND CONCENTRATION IS EXCEEDED

The following procedure describes a t-test to determine significance if an SSO for a chemical with a high natural background concentration is exceeded. If an SSO is simply defined as the mean concentration of a chemical at the upstream (or reference) site (no input due to human activities), significant differences in average concentrations between the upstream and downstream sites ($X_d - X_u$) can be analyzed with a *t* test, provided that samples are normally distributed and have equal variances (Zar, 1999). The *t* statistic is

$$t = \frac{\overline{X}_S - \overline{X}_R}{s_{\overline{X}d - \overline{X}_u}}$$

where:

$$s_{\overline{X}d - \overline{X}u} = \sqrt{\frac{s_p^{\ 2}}{n_d} + \frac{s_p^{\ 2}}{n_u}}$$

The pooled variance (s^2_p) equals the sum of squares (SS) from the study and reference sites, divided by the sum of their degrees of freedom ($v_d + v_u = n - 2$).

$$s_p^2 = \frac{\mathbf{SS}_d + \mathbf{SS}_u}{v_d + v_u} \qquad \text{and} \qquad \mathbf{SS} = \sum_{i=1}^{n} (X_i - \overline{X})^2$$

We reject the null hypothesis H_0: $\mu_d \# \mu_u$ if $t \exists t_{\forall(1),v.}$ when samples from the upstream and downstream sites have unequal variances. Welch's approximate *t* offers a reliable alternative test. The critical value for *t*' is Student's $t_{\forall(1)}$ with *v*' degrees of freedom, where

$$t' = \frac{\overline{X}_d - \overline{X}_u}{\sqrt{\frac{s_d^2}{n_d} + \frac{s_u^2}{n_u}}} \qquad \text{and} \qquad v' = \frac{\left(\frac{s_d^2}{n_d} + \frac{s_u^2}{n_u}\right)^2}{\frac{\left(\frac{s_d^2}{n_d}\right)^2}{n_d - 1} + \frac{\left(\frac{s_u^2}{n_u}\right)^2}{n_u - 1}}$$

The calculated v' degrees of freedom will generally include decimals lower than one, but only the next smaller integer should be used.

When concentrations of a chemical in the reference site are compared against multiple locations, the overall probability of rejecting a true null hypothesis (Type I error) may be far greater than the reported error rate for each individual test. For instance, if we perform twenty tests at the 0.05 level of significance, the probability of at least one Type I error is

$$1 - (1-0.05)^{20} = 0.64$$

Dunnett's test (1955) is specifically designed to adjust the error rate for multiple comparisons between the mean of a reference site and the means of (k-1) study areas. The test statistic is

$$q = \frac{X_u - X_d}{\mathbf{SE}}$$

where:

$$\mathbf{SE} = \sqrt{s^2\left(\frac{1}{n_u} + \frac{1}{n_d}\right)}$$

s^2 is an estimate of the error mean square (from an analysis of variance to test the null hypothesis H_0:

$$\mu_u = \mu_{d_1} = \mu_{d_2} = \ldots = \mu_{d_{k-1}}$$

Critical values for the Dunnett's statistic ($q'_{\forall(1),v,k}$) are available from published tables (e.g., Zar 1999). The number of samples from the reference site should be greater than the number of samples from any of the study areas. The recommended, optimal sample size for the reference site should be slightly less than

$$\sqrt{k-1}$$

times the sample size for each of the study areas.

We use simulated data on concentrations of aluminum at one reference and three study sites to illustrate an application of Dunnett's test. Aluminum concentrations are expressed in (mg/L)

Reference Site

0.6793	0.5138	0.5168	0.1167	0.5241	0.5955	0.6112	0.5346	0.4837	0.7029
0.6900	0.7090	0.7246	0.6152	0.6192	0.4560	0.5283	0.3745	0.1521	0.2344
0.7703	0.6988	0.6569	0.4905	0.5114	0.5195	0.5560	0.6615	0.2793	0.5210
0.6760	0.8210	0.7340	0.7960	0.8315	0.6755	0.4892	0.5408	0.3602	0.5258
0.5192	0.5281	0.8750	0.9434	0.8414					

<u>Site 1</u>

0.6960	0.6437	0.6424	0.4495	0.5783	0.4619	0.7810	0.5243	0.5479	0.4735
0.6910	0.4022	0.4355	0.3525	0.4347	0.2777	0.4713	1.0165	0.3342	0.8989
0.4965	0.3570	0.6228	0.7117	0.7885	0.7472	0.6538	0.2597	0.5368	0.4852

<u>Site 2</u>

0.6202	0.7330	0.4285	0.7584	0.8533	0.6029	0.7708	1.0277	0.4256	0.6523
0.7853	0.3032	0.7133	0.6175	0.8988	0.4975	0.4835	0.7949	0.6894	0.6831
0.4077	0.5171	0.6178	0.6930	0.6399	0.4603	0.4202	0.6798	0.5955	0.5000

<u>Site 3</u>

0.5092	0.7135	0.5944	1.0061	0.6414	0.3645	0.5905	0.5620	0.5237	0.6052
0.3857	0.8539	0.8968	0.5050	0.6330	0.8206	0.5120	0.7247	0.8545	0.3682
0.4925	0.5490	0.7268	0.9285	0.4859	0.3341	0.5429	0.8066	0.5864	0.7111

The first step of the analysis is to test for normality of the data, a critical assumption of Dunnett's test. Departures from the normal distribution were assessed with the Shapiro and Wilk test (1965). The assumption of normality is rejected if the computed value of the test statistic (W) is less than the critical value ($W_{\forall}$). The error rate (∀) for these multiple comparisons can be adjusted with the sequential method proposed by Rice (1990), but such procedure was not necessary because all computed Ws (Table 1) were greater than 0.1. The Shapiro and Wilk tests were performed in S-Plus 6 (Insightful Corporation 2001). Next, we used the Dunnett-type test of Levy (1975) to compare the variance (s^2) of each study site with variance of the reference area. The test statistic is

$$q = \frac{\ln s_u^2 - \ln s_d^2}{\mathbf{SE}}$$

where:

$$\mathbf{SE} = \sqrt{\left(\frac{2}{v_u}\right) + \left(\frac{2}{v_d}\right)}$$

Variance in aluminum concentrations at the three study sites were not significantly different from variance in aluminum concentration at the reference area ($q'_{0.05(2),131,4}$ □ 2.35, Table 1). Therefore, the assumption of homogeneity of variances was satisfied for all comparisons. Differences in mean concentration of aluminum between the reference area and sites 1 and 2 were not significant (0.346, 1.064 < $q'_{0.05(1),131,4}$ □ 2.08, Table 1). The average concentration of aluminum at site 3 was significantly higher than the average concentration at the reference site (3.351 > $q'_{0.01(1),131,4}$ □ 2.97).

TABLE 1. SUMMARY OF DATA AND RESULTS OF DUNNETT'S TEST (MEAN | Q|) COMPARING THE AVERAGE ALUMINUM CONCENTRATION IN A REFERENCE AREA WITH EACH OF THE THREE TARGET SITES

	n	Mean	SD	W	Variance |*q*|	Mean |*q*|
Reference	45	0.5823	0.1795	0.96	-	-
Site 1	30	0.5975	0.1651	0.96	0.493	0.346
Site 2	30	0.6290	0.1641	0.98	0.528	1.064
Site 3	30	0.7294	0.2315	0.96	1.506	3.351

When the data significantly departs from a normal distribution, the nonparametric, one-tailed Mann-Whitney test is recommended. It tests for differences in ranks between the reference and study sites, instead of actual measurements of aluminum concentration. There is also a nonparametric procedure to adjust the error rate for multiple comparisons (see Zar 1999: 225).

References

Dunnett, C.W. 1955. A multiple comparison procedure for comparing several treatments with a control. *Journal of the American Statistical Association* **50**:1096-1121.

Insightful Corporation. 2001. *S-Plus 6.0 for Windows, Professional Release 2*. Seattle, Washington.

Rice, W.R. 1990. A consensus combined P-value test and the family-wide significance of component tests. *Biometrics* 46:303-308.

Shapiro, S.S. and M.B. Wilk. 1965. An analysis of variance test for normality (complete samples). *Biometrika* **52**:591-561.

Zar, J.H. 1999. *Biostatistical Analysis*. Prentice Hall. Upper Saddle River, New Jersey.